ChinaXiv合作期刊

贪心非对称深度有监督哈希图像检索方法 *

赵昕昕，李 阳，苗 壮†，王家宝，张 睿

(陆军工程大学 指挥控制工程学院，南京 210007)

摘 要：近年来，深度有监督哈希检索方法已成功应用于众多图像检索系统中。但现有方法仍然存在一些不足：一是大部分深度哈希学习方法都采用对称策略来训练网络，但该策略训练通常比较耗时，难以用于大规模哈希学习过程；二是哈希学习过程中存在离散优化问题，现有方法将该问题进行松弛，但难以保证得到最优解。为解决上述问题，提出了一种贪心非对称深度有监督哈希图像检索方法，该方法将贪心算法和非对称策略的优势充分结合，进一步提高了哈希检索性能。在两个常用数据集上与 17 种先进方法进行比较。在 CIFAR-10 数据集上 48 比特条件下，与性能最好的方法相比 mAP 提高 1.3%；在 NUS-WIDE 数据集上所有比特下，mAP 平均提高 2.3%。在两个数据集上的实验结果表明，所提方法可以进一步提高哈希检索性能。

关键词：非对称策略；贪心算法；有监督哈希；图像检索

中图分类号：TP391 doi: 10.19734/j.issn.1001-3695.2022.03.0076

Greedy-asymmetric deep supervised hashing for image retrieval

Zhao Xinxin, Li Yang, Miao Zhuang†, Wang Jiabao, Zhang Rui

(*Command & Control Engineering College*, *Army Engineering University of PLA*, *Nanjing* 210007, *China*)

Abstract: In recent years, the deep supervised hash retrieval method have been successfully applied to many image retrieval systems. However, the existing methods still have some shortcomings: First, most of the deep hash learning methods use symmetric strategies to train the network, but the training of this strategy is usually time-consuming and difficult to be used in the large-scale hash learning process; Second, there is a discrete optimization problem in the hash learning process. Existing methods relax this problem and it is difficult to guarantee the optimal solution. In order to solve the above problems, this paper proposes a greedy-asymmetric deep supervised hashing method for image retrieval, which fully combines the advantages of the greedy algorithm and asymmetric strategy to further improve the hash retrieval performance. This article compares 17 state-of-the-art methods on two commonly used datasets. Compared with the state-of-the-art methods, the propoed method increases the mAP in 48-bits setting by 1.3% on CIFAR-10 dataset. And on NUS-WIDE dataset, increases the mAP in all-bits setting by increased 2.3% on average. The experimental results show that the propoed method can further improve the performance of hash retrieval.

Key words: asymmetric strategy; greedy algorithm; supervised hashing; image retrieval

0 引言

随着图像数据的爆炸性增长，如何从海量数据中寻找所需要的信息已经成为一个至关重要的问题。在大规模图像检索中，由于哈希检索方法[1]在计算和存储方面具有高效性[2]，因此基于哈希特征的近似最近邻搜索方法[3]受到了广泛关注。

近年来，许多哈希检索方法已经被提出，它们可以分为数据无关[4]和数据相关方法[5]。数据无关方法主要是依靠随机映射来构造哈希函数，而哈希函数构造过程通常不依赖于训练数据，因此该类方法的检索精度较低[6]。数据相关方法利用各种机器学习技术来学习哈希函数。与数据无关方法相比，数据相关方法能够利用更短的哈希码达到更高的精度[7]。因此，数据相关方法比数据无关方法得到了更加广泛的应用。

由于深度神经网络在图像分类[8]、目标检测[9]、人脸识别[10]等众多任务[11]中的成功应用，研究人员试图通过深度学习来解决哈希问题，并提出了深度有监督哈希学习方法。深度有监督哈希学习方法是通过深度神经网络提取出图像特征，并同时进行哈希学习的一类方法[12, 13]。该类方法将特征提取和哈希学习集成到同一个端到端框架中[14]，使得哈希编码的检索性能相对于传统方法得到了大幅提升。

尽管深度有监督哈希学习方法已经取得了巨大的进步，但依然存在许多不足[15, 16]。现有的深度有监督哈希学习方法大部分在训练过程中采用对称策略，例如，CNNH[17] (Convolutional Neural Networks Hashing)、DPSH[18] (Deep Pairwise Supervised Hashing)、DHN[19] (Deep Hashing Network)等。对称策略是指使用相同的深度神经网络对查询样本点和数据集合样本点进行深度特征提取，再使用相同的深度哈希函数为查询样本点和数据集合样本点生成哈希码。采用对称策略的哈希学习方法的训练通常比较耗时，这使得它们很难有效地利用监督信息来处理大规模数据集合。例如，DPSH 方法[18]的存储和计算成本是 $O(n^2)$，其中 n 是数据集合样本点的数量；DTSH[20]方法(Deep Supervised Hashing with Triplet)的训练成本甚至更高。由于采用对称策略训练网络比较耗时，所以为了使训练切实可行，大多数现有的方法只从整个数据集合中抽取一个子集来构建哈希函数学习的训练集，而其余数据集合样本点则被丢弃。采用以上策略虽然能够加速网络

收稿日期：2022-03-03；**修回日期：**2022-04-17 **基金项目：**国家自然科学基金资助项目(61806220)；江苏省自然科学基金资助项目(BK20200581)

作者简介：赵昕昕(1997-)，女，河南平顶山人，硕士研究生，主要研究方向为机器学习、图像检索；李阳(1984-)，男，河北廊坊人，副教授，硕导，博士，主要研究方向为机器视觉、机器学习；苗壮(1976-)，男(通信作者)，辽宁辽阳人，副教授，硕导，博士，主要研究方向为人工智能(emiao_beyond@163.com)；王家宝(1985-)，男，安徽肥西人，讲师，博士，主要研究方向为模式识别、机器学习；张睿(1977-)，男，山东威海人，教授，博士，主要研究方向为信息融合.

chinaXiv:202205.00069v1

chinaXiv:202205.00069v1

训练，但是会出现监督信息利用不充分问题[21]。

在早期关于非对称哈希[22, 23]的研究工作中，它们使用非对称距离度量来保持图像之间的相似度，其中训练集和查询集的二进制哈希码是由需要学习的相同的哈希函数生成的。受上述非对称思想启发，后续工作中又提出了一些深度非对称哈希方法，如 Deep Asymmetric Pairwise Hashing[24]、Nonlinear Asymmetric Multi-valued Hashing[25]、Collaborative learning for extremely low bit Asymmetric Hashing[26]、Deep Asymmetric Hashing with Dual Semantic Regression and Class Structure Quantization[27]和基于非对称监督深度离散哈希的图像检索[28]等。其中 ADSH[21]方法采用非对称方式进行深度哈希学习。非对称方式是指特征提取仅仅只对查询样本点执行，而不对数据集合样本点执行。在深度哈希函数学习过程中只学习查询样本点的哈希码，而数据集合样本点的哈希码则是直接学习得到。该方法所提的非对称方式能够有效解决上述监督信息利用不充分问题。

虽然 ADSH 方法通过非对称方式能够充分利用监督信息并实现高效地训练网络，但由于该方法在哈希码学习过程中，需要使用 *sign* 函数(*sign* 函数不可导)将深度特征映射为二值化哈希码，进而使得该优化问题变成 NP 难问题。ADSH 方法在训练阶段使用 *tanh* 函数替换 *sign* 函数进行松弛[29]，并在损失函数中添加一个惩罚项来生成尽可能离散的特征[18, 30]，然后在测试阶段再使用 *sign* 函数来获得真正的二进制哈希码。虽然这样能够训练网络，但该方法会引入量化误差。为了更好求解离散优化问题，Greedy Hash[31]方法利用贪心原理在正向传播中严格使用 *sign* 函数保持对网络输出的离散约束，而在反向传播中，哈希层梯度被完整地传送到前层，进一步避免了梯度消失。

在 ADSH 和 Greedy Hash 方法的启发下，本文提出一种贪心非对称深度有监督哈希图像检索方法。该方法将贪心算法和非对称策略同时应用到哈希函数的学习过程中，使得该学习过程即可充分利用监督信息又能更好求解离散优化问题。

具体地，本文的主要贡献包括以下 3 个方面：

a)提出了一种贪心非对称深度有监督哈希图像检索方法，该方法将贪心算法和非对称策略的优势充分结合，进一步提高了哈希检索性能。

b)设计了一种贪心非对称成对损失，该损失函数包括贪心损失和非对称成对损失两部分。贪心损失通过在前传过程中保持对网络输出的离散约束，在反传过程中将哈希层的梯度完全传送到网络输出层，进一步解决离散优化问题。非对称成对损失通过采用非对称策略对查询样本点和数据集合样本点使用不同策略进行学习，进一步提高了哈希码学习效率。

c)在两个数据集上的实验结果表明，贪心非对称深度有监督哈希图像检索方法可以进一步提高哈希检索性能。

1　特征提取

图 1 展示了贪心非对称深度有监督哈希图像检索方法的模型结构。该模型结构主要包括两部分：特征提取部分和损失函数部分。特征提取部分的主要作用是通过一个骨干网提取出图像的深度特征。损失函数部分主要作用是通过哈希学习将图像深度特征映射成哈希码。如图 1 所示，训练图像首先在特征提取阶段得到图像的深度特征，其次通过哈希层得到图像的二进制哈希码，最后通过所设计的损失函数指导网络训练生成能够保留图像相似性的哈希码。该方法将两部分融入到同一个端到端的结构中，使得模型在训练过程中各部分可实现相互反馈，实现了更加鲁棒的哈希编码。

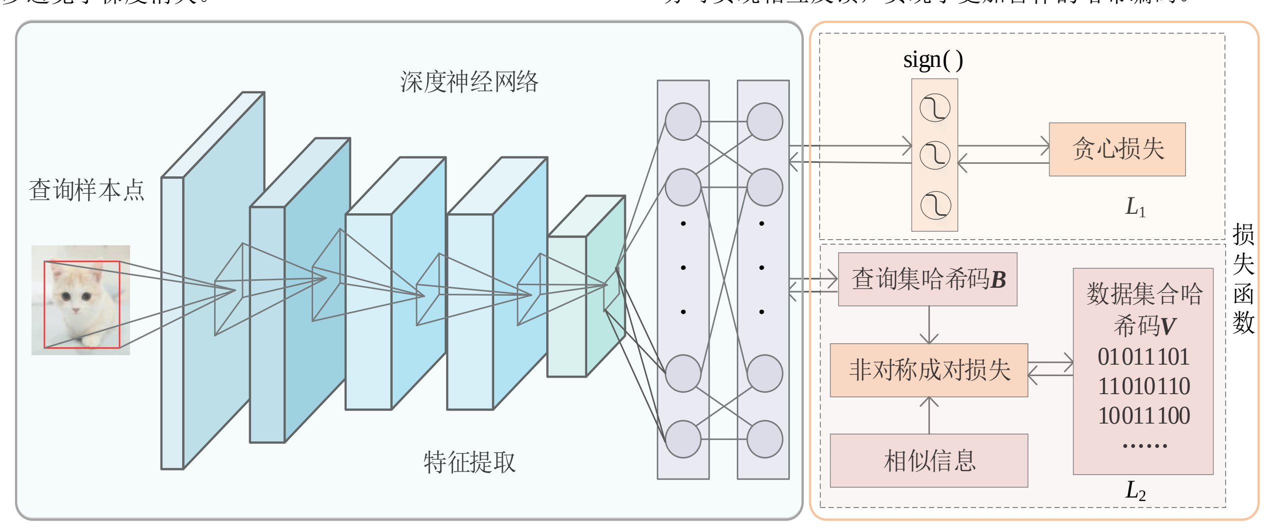


图 1　本文方法总体架构

Fig. 1　The overall framework of the proposed method

值得注意的是，本文设计的损失函数包括两部分：贪心损失和非对称成对损失。如图 1 中贪心损失 L_1 所示，得到图像的哈希特征后直接使用 *sign* 函数将哈希特征映射成二进制哈希码，再通过贪心原理解决离散优化问题，具体实现过程如 2.1 节所示。如图 1 中非对称成对损失 L_2 所示，该部分同时使用查询图像与数据集合图像的监督信息对网络进行训练，并在训练过程中能够学习得到数据集合图像的二进制哈希码。该部分具体实现过程如 2.2 节所示。

如图 1 中特征提取阶段所示，为了提取图像的深度特征，本文采用 AlexNet[8]作为骨干网进行特征提取。AlexNet 骨干网包括 5 个卷积层和 3 个全连接层(结构如表 1 所示)。为了得到图像的哈希特征，本文在 AlexNet 骨干网的 FC3 层后添加一个哈希编码层，它可以将深度特征映射到 $\boldsymbol{R}^c$ 空间中，c 为哈希码长度。

表 1　AlexNet 神经网络结构

Tab. 1　The network structure of alexnet

Type	Filter size/stride	Output size
Conv1	11×11/4	55×55×96
Conv2	5×5/1	27×27×256
Conv3	3×3/1	13×13×384
Conv4	3×3/1	13×13×384
Conv5	3×3/1	13×13×256
FC1	4096	4096
FC2	4096	4096
FC3	1000	1000

骨干网的输入为图像 $x_i \in \boldsymbol{X}$，$\boldsymbol{X}=\{x_i\}_{i=1}^{m}$ 表示查询图像集合，

ChinaXiv合作期刊

该数据集中共包含 m 个样本。图像经过骨干网提取得到哈希特征 $\boldsymbol{h}_i$：

$$\boldsymbol{h}_i = f(x_i, \theta) \tag{1}$$

其中 f 表示骨干网函数，θ 表示骨干网的参数。

2　损失函数

如图 1 中损失函数阶段所示，为学习得到更优的哈希码，设计了一个损失函数，可表示为：$L = L_1 + \lambda L_2$，其中 L_1 为贪心损失，L_2 为非对称成对损失，λ 为超参数。贪心损失 L_1 能够解决优化过程中梯度消失问题[30]；非对称成对损失能够充分利用数据集合标签信息，并高效训练网络[21]。

2.1　贪心损失

为了得到图像的哈希编码，通常会在哈希编码层后面使用 *sign* 函数将深度特征映射为二值化哈希码。但由于 *sign* 函数不可导，会使优化过程变为 NP 难问题。传统方法使用 *tanh* 或 *sigmoid* 函数进行松弛[29]，这样虽然能够训练网络，但会产生次优解。为更好解决这个问题，本文提出利用贪心算法，在正向传播中严格使用 *sign* 函数保持对网络输出的离散约束；而在反向传播中，哈希层梯度被完整地传送到前层，避免了梯度消失，有效地解决了离散优化问题。

贪心损失的核心问题是如何利用贪心算法来解决离散优化问题。因此，该损失关注离散优化问题 $\min L_1(\boldsymbol{b}_i)$，其中 $\boldsymbol{b}_i = sign(\boldsymbol{h}_i) \in \{-1, 1\}^{1\times c}$ 表示 x_i 的哈希码。具体地，本文的贪心损失采用交叉熵损失形式。

在离散优化过程中，如果完全忽略离散约束 $\boldsymbol{b}_i \in \{-1, 1\}^{1\times c}$，利用梯度下降算法可以得到第$(t+1)$次迭代的 b_i，即

$$\boldsymbol{b}_i^{t+1} = \boldsymbol{b}_i^t - l_r * \frac{\partial L_1}{\partial \boldsymbol{b}_i^t} \tag{2}$$

其中 l_r 代表学习率。然而，使用式(2)得到的解不满足 $\boldsymbol{b}_i^{t+1} \in \{-1, 1\}^{1\times c}$。若不考虑离散约束，则由式(2)所得到的解为连续最优解。贪心原理认为离连续最优解最近的离散点如式(3)，就是所希望得到的离散最优解。

$$\boldsymbol{b}_i^{t+1} = sign(\boldsymbol{b}_i^t - l_r * \frac{\partial L_1}{\partial \boldsymbol{b}_i^t}) \tag{3}$$

式(3)可拆分为前向传播和反向传播两步来实现。前向传播过程如式(4)所示，它通过新的哈希层在正向传播中使用 *sign* 函数实现：

$$\boldsymbol{b}_i^{t+1} = sign(\boldsymbol{h}_i^{t+1}) \tag{4}$$

反向传播过程中，本文在贪心损失中添加一个惩罚项 $\|\boldsymbol{h}_i - sign(\boldsymbol{h}_i)\|_p^p$，并且使得该惩罚项尽可能地接近零。再由 $\boldsymbol{b}_i^t = sign(\boldsymbol{h}_i^t)$ 可以得到：

$$\begin{aligned}\boldsymbol{h}_i^{t+1} = \boldsymbol{h}_i^t - l_r * \frac{\partial L_1}{\partial \boldsymbol{h}_i^t} = (\boldsymbol{h}_i^t - \boldsymbol{b}_i^t) + \boldsymbol{b}_i^t - l_r * \frac{\partial L_1}{\partial \boldsymbol{h}_i^t} = \\ (\boldsymbol{h}_i^t - sign(\boldsymbol{h}_i^t)) + \boldsymbol{b}_i^t - l_r * \frac{\partial L_1}{\partial \boldsymbol{h}_i^t} \approx \boldsymbol{b}_i^t - l_r * \frac{\partial L_1}{\partial \boldsymbol{h}_i^t}\end{aligned} \tag{5}$$

令

$$\frac{\partial L_1}{\partial \boldsymbol{h}_i^t} = \frac{\partial L_1}{\partial \boldsymbol{b}_i^t} \tag{6}$$

即可得到：

$$\boldsymbol{h}_i^{t+1} = \boldsymbol{b}_i^t - l_r * \frac{\partial L_1}{\partial \boldsymbol{b}_i^t} \tag{7}$$

式(6)表示，在反向传播过程中，贪心原理能够将哈希层梯度完全传送到网络输出层。通过分别实现正向传播和反向传播过程，本文的贪心损失有效地解决了离散优化问题，且获得精确的哈希码。

2.2　非对称成对损失

非对称成对损失 L_2 的作用是：在训练过程中，采用非对称策略训练网络。这样不仅能够充分利用数据集合的监督信息，而且也可以高效训练网络。所谓非对称策略是指，采用不同的方式来处理查询图像和数据集合图像。对于查询图像，通过骨干网进行深度特征提取，再使用深度哈希函数生成查询图像的哈希码；而对于数据集合图像，它的哈希码则是直接学习得到。

为得到能够保留查询图像与数据集合图像之间相似性的哈希码，一种常见的方法[29]是最小化训练图像与数据集合图像的监督信息相似性和哈希码内积之间的相似性：

$$\min_{\boldsymbol{B},\boldsymbol{V}} L_2(\boldsymbol{B}, \boldsymbol{V}) = \sum_{i=1}^{m}\sum_{j=1}^{n}[\boldsymbol{b}_i^T \boldsymbol{v}_j - c\boldsymbol{S}_{ij}]^2 \tag{8}$$

其中 $\boldsymbol{B} = \{\boldsymbol{b}_i\}_{i=1}^m \in \{-1, 1\}^{m\times c}$，表示训练图像集合通过哈希函数所学习到的哈希码集合。$\boldsymbol{V} = \{\boldsymbol{v}_j\}_{j=1}^n \in \{-1, 1\}^{n\times c}$ 表示直接学习得到的数据集合图像对应的哈希码集合，该集合中包含 n 个样本点。$\boldsymbol{S}_{ij}$ 表示相似矩阵，如果两张图像相似，$\boldsymbol{S}_{ij} = 1$，否则 $\boldsymbol{S}_{ij} = -1$。

由 $\boldsymbol{h}_i = f(x_i, \theta)$ 和 $\boldsymbol{b}_i = sign(\boldsymbol{h}_i)$，式(8)可变换为

$$\min_{\theta,\boldsymbol{V}} L_2(\theta, \boldsymbol{V}) = \sum_{i=1}^{m}\sum_{j=1}^{n}[sign(f(x_i, \theta))^T \boldsymbol{v}_j - c\boldsymbol{S}_{ij}]^2 \tag{9}$$

但由于式(9)中存在 *sign* 函数(*sign* 函数不可导)，因此参数 θ 的梯度不能直接进行反向传播。本文使用 *tanh* 函数替代 *sign* 函数进行松弛，式(9)表示如下：

$$\min_{\theta,\boldsymbol{V}} L_2(\theta, \boldsymbol{V}) = \sum_{i=1}^{m}\sum_{j=1}^{n}[tanh(f(x_i, \theta))^T \boldsymbol{v}_j - c\boldsymbol{S}_{ij}]^2 \tag{10}$$

在实际应用中，如果只给定一个数据集合 $\boldsymbol{Z}$ 而没有指明查询集 $\boldsymbol{X}$，那么可以从 $\boldsymbol{Z}$ 中随机采样 m 个数据作为查询集。即 $\boldsymbol{X} = \boldsymbol{Z}^{\boldsymbol{\Omega}}$，其中 $\boldsymbol{Z}^{\boldsymbol{\Omega}} = \{z_1, z_2, \cdots, z_m\}$ 表示从数据集合中随机采样的数据集合，$\boldsymbol{\Omega} = \{i_1, i_2, \cdots, i_m\}$ 表示采样得到数据的索引集。本文令 $\boldsymbol{\Gamma}$ 表示数据集合的所有索引值，则 $\boldsymbol{\Omega} \subseteq \boldsymbol{\Gamma}$。同样可以得到采样数据集对应的相似矩阵 $\boldsymbol{S}^{\boldsymbol{\Omega}}$，令 $\boldsymbol{S} = \boldsymbol{S}^{\boldsymbol{\Omega}}$。则式(10)可重写为

$$\min_{\theta,\boldsymbol{V}} L_2(\theta, \boldsymbol{V}) = \sum_{i\in\boldsymbol{\Omega}}\sum_{j\in\boldsymbol{\Gamma}}[tanh(f(z_i, \theta))^T \boldsymbol{v}_j - c\boldsymbol{S}_{ij}]^2 \tag{11}$$

为保持 $\boldsymbol{v}_i$ 和 $tanh(f(z_i, \theta))$ 尽可能接近，在式(11)中添加一个额外约束是合理的。由于 $tanh(f(z_i, \theta))$ 是 z_i 的哈希码的近似值，因此式(11)可重写为

$$\begin{aligned}\min_{\theta,\boldsymbol{V}} L_2(\theta, \boldsymbol{V}) = \sum_{i\in\boldsymbol{\Omega}}\sum_{j\in\boldsymbol{\Gamma}}[tanh(f(z_i, \theta))^T \boldsymbol{v}_j - c\boldsymbol{S}_{ij}]^2 + \\ \gamma\sum_{i\in\boldsymbol{\Omega}}[\boldsymbol{v}_i - tanh(f(z_i, \theta))]^2\end{aligned} \tag{12}$$

其中 γ 为超参数。

本文采用交替优化策略来学习式(12)中的参数 θ 和 $\boldsymbol{V}$，该方法也可适用于式(10)。具体地，在每次迭代中只学习一个参数而其他参数固定，这个过程重复多次迭代。交替优化策略更新参数的具体过程可参考 ADSH[21]方法中学习算法部分。

3　实验

为验证所提出方法的有效性，本文在两个常用公开数据集上进行了测试，并同 17 种方法进行了比较。

3.1　数据集

a)CIFAR-10 数据集。CIFAR-10 数据集[32]共包含 10 个类，每类包含 6000 个样本，总共有 60000 张彩色图像，图像大小为 32×32。对于 CIFAR-10 数据集，如果两幅图像标签相同，那么将两张图像视为相似对。

b)NUS-WIDE 数据集。NUS-WIDE 数据集[33]由 269648 张带标签的网络图像组成。它是一个多标签数据集，其中每个图像都包含多张标签。本文只选择了 10 个最常见的类的图像。对于 NUS-WIDE 数据集，如果两张图像至少共享一个公共标签，它们将被定义为相似对。

3.2　实验设置

本实验的模型是在 PyTorch 框架下实现的。min-batch 的大小设置为 128，并使用 Adam[34]作为优化器，权重衰减设置

chinaXiv:202205.00069v1

ChinaXiv合作期刊

chinaXiv:202205.00069v1

为 0.0005，最大迭代次数设置为 50。在数据划分上，对于 CIFAR-10 数据集，随机从每类中抽取 200 张，总共抽取 2000 张图像作为训练集；从数据集合中每类抽取 100 张，总共 1000 张图像作为查询集。除去 1000 张查询集之外的 59000 张图像作为数据集合。

对于 NUS-WIDE 数据集，每次从数据集合中随机抽取 2000 张作为训练集，随机抽取 1000 张作为查询集，除去查询集之外的所有图像作为数据集合。在评价指标中，本文选取图像检索中最常用的评价指标：mAP (mean Average Precision)和 PR (Precision-Recall)。

在深度哈希方法中，由于不同深度神经网络具有不同的特征提取能力，在选取不同深度神经网络作为骨干网时检索性能会有较大变化。为便于比较，本实验采用在 ImageNet 数据集[35]上预训练的 AlexNet 模型作为骨干网。

3.3　实验结果与分析

3.3.1 检索结果对比

本实验在图像数据集 CIFAR-10 和 NUS-WIDE 上进行，为评估本文贪心非对称深度有监督哈希的性能，本文选取 17 种方法进行比较，其中包括 7 种传统哈希学习方法和 10 种深度哈希学习方法。传统哈希学习方法中包括无监督哈希学习方法：ITQ[2]，有监督哈希学习方法：Lin:Lin[35]、LFH[37]、FastH[38]、SDH[39]、COSDISH[40]、KADGH[41]。深度哈希学习方法包括：DSH[1]、DSDH[15]、DHN[19]、DPSH[18]、DTSH[20]、Greedy Hash[31]、ADSH[21]、SDNMSH[42]、ASDDH[27] 和 TransHash[43]。

在 CIFAR-10 和 NUS-WIDE 上的图像检索 mAP 精度如表 2 所示，加粗字体表示最优值，下划线表示次优值，*表示本文复现结果。从表 2 中可以看出，在大多数情况下，有监督哈希学习方法性能都要优于无监督哈希学习方法，而深度哈希学习方法优于传统哈希学习方法。同时，本文方法在所有长度哈希码上的检索性能均显著优于其他方法。这是因为所提出的贪心非对称损失能够更好地保留图像特征信息，从而提高哈希检索性能。在 CIFAR-10 数据集上，本文方法在 48 比特条件下的性能比 ADSH 方法提高了 3.0%，与 TransHash 方法相比提高了 1.4%。在 NUS-WIDE 数据集上，所提方法在不同比特下检索性能平均提高了 2.3%。从实验结果可以看出，本文方法在这两个常用数据集上的性能均较好，尤其是在单标签数据集 CIFAR-10 上检索性能提升更加明显。

表 2　CIFAR-10 和 NUS-WIDE 上图像检索 mAP 精度对比

Tab. 2　Comparison of map accuracy on CIFAR-10 and NUS-WIDE

方法	CIFAR-10				NUS-WIDE			
	12 位	24 位	32 位	48 位	12 位	24 位	32 位	48 位
ITQ[2]	0.2619	0.2754	0.2861	0.2941	0.7143	0.7361	0.7457	0.7553
Lin: Lin[36]	0.6099	0.6312	0.6079	0.6013	0.5556	0.5704	0.5627	0.5555
LFH[37]	0.4178	0.5738	0.6414	0.6927	0.7116	0.7681	0.7949	0.8135
FastH[38]	0.5971	0.6632	0.6847	0.7020	0.7267	0.7692	0.7817	0.8037
SDH[39]	0.4539	0.6334	0.6514	0.6603	0.7646	0.7998	0.8017	0.8124
COSDISH[40]	0.5831	0.6614	0.6802	0.7016	0.6425	0.7406	0.7843	0.7964
KADGH[41]	0.6134	0.6607	0.6701	0.6829	--	--	--	--
Greedy Hash[31]	0.7740	0.7950	0.8100	0.8200	--	--	--	--
SDNMSH[42]	--	0.8068	0.8106	0.8135	--	0.6169	0.6234	0.6289
DSDH[15]	0.7400	0.7860	0.8010	0.8200	0.5930	0.5930	0.6200	0.6350
DPSH[18]	0.6818	0.7204	0.7341	0.7464	0.7941	0.8249	0.8351	0.8442
DTSH[20]	0.7100	0.7500	0.7650	0.7740	0.7730	0.8080	0.8120	0.8240
DSH[1]	0.6441	0.7421	0.7703	0.7992	0.7125	0.7313	0.7401	0.7485
DHN[19]	0.6805	0.7213	0.7233	0.7332	0.7719	0.8013	0.8051	0.8146
ASDDH[27]	0.7360	0.7710	0.7810	0.7850	0.8340	0.8510	0.8680	0.8740
TransHash[43]	--	--	0.9108	0.9141	--	--	0.7393	0.7532
ADSH*[21]	0.8950	0.9010	0.9030	0.8970	0.8690	0.8990	0.9050	0.9110
本文方法	**0.9130**	**0.9150**	**0.9160**	**0.9270**	**0.8940**	**0.9180**	**0.9350**	**0.9300**

为了进一步说明所提方法的优越性，本文在 CIFAR-10 和 NUS-WIDE 数据集上 12 比特条件下分别绘制了 PR 曲线，如图 2、3 所示。PR 曲线与横坐标轴所围面积越大，则表示该方法性能越好。从图 2、3 各方法的 PR 曲线与横坐标轴所围面积可以看出，贪心非对称深度有监督哈希图像检索方法的性能明显优于其他所有方法。

3.3.2 超参数分析

贪心非对称损失为 $L = L_1 + \lambda L_2$，其中 L_1 为贪心损失，L_2 为非对称成对损失，λ 为超参数。为分析 λ 对于检索性能的影响，本实验在 CIFAR-10 和 NUS-WIDE 数据集上进行参数分析。图 4、5 显示了本实验分别在 12 比特和 48 比特下 λ 取不同值(0.1、1、10、100、1000)时的检索精度。从图 4、5 中可以看出，超参数 λ 的取值对两个数据集的检索结果影响并不明显。图 4 显示：当 $0.1 < \lambda < 1$ 时，CIFAR-10 数据集能够达到较好的检索性能。$\lambda = 1$ 时，CIFAR-10 在不同比特下的检索结果均能达到最好；图 5 显示：当 $0.1 < \lambda < 10$ 时，NUS-WIDE 数据集能达到较好的检索性能。$\lambda = 1$ 时，NUS-WIDE 在不同比特下的检索结果均能最好。综上所述，本文超参数 λ 取值为 1。

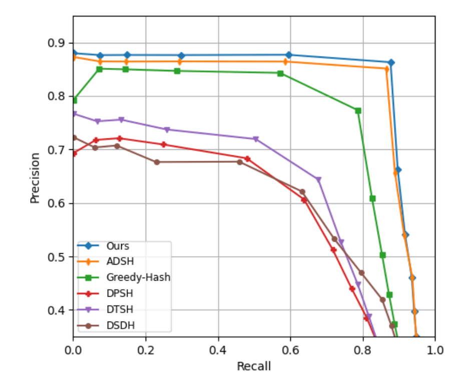


图 2　CIFAR-10 PR 曲线

Fig. 2　PR curve of CIFAR-10

ChinaXiv合作期刊

chinaXiv:202205.00069v1

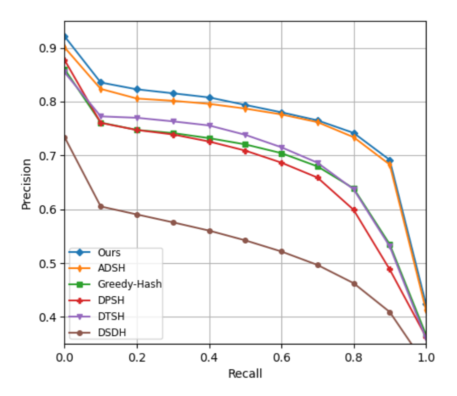


图 3　NUS-WIDE PR 曲线

Fig. 3　PR curve of NUS-WIDE

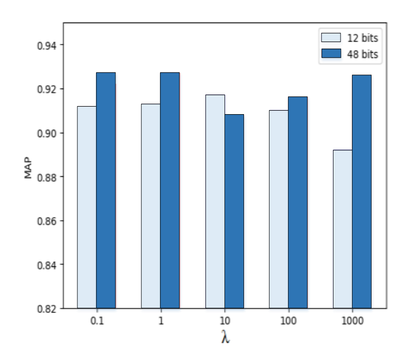


图 4　λ 对 CIFAR-10 检索精度的影响

Fig. 4　The retrieval accuracy with different λ on CIFAR-10

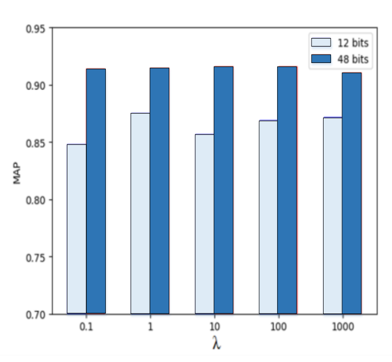


图 5　λ 对 NUS-WIDE 检索精度的影响

Fig. 5　The retrieval accuracy with different λ on NUS-WIDE

3.3.3 检索结果可视化

为了进一步说明本文方法的检索效果，图 6 中给出了对 CIFAR-10 数据集每类图像查询结果(TOP 10 检索结果)的可视化。从图 6 可以看出，CIFAR-10 数据集的 TOP 10 平均检索精度可以达到 98%。进一步验证了本文方法在单标签数据集上优异的检索性能。



图 6　CIFAR-10 的 TOP 10 检索结果

Fig. 6　Top 10 retrieval results on CIFAR-10

4　结束语

针对哈希函数学习过程中的离散优化问题和监督信息利用不充分问题，提出了贪心非对称深度有监督哈希图像检索方法。该方法将贪心算法和非对称策略同时应用到哈希学习过程中，这样不仅能够将哈希层梯度完全传送到网络输出层来解决离散优化问题，而且能够充分利用监督信息高效地训练网络。所提方法与 17 种方法在两个公开数据集上进行对比实验，验证了该方法的有效性。虽然贪心非对称深度有监督哈希图像检索方法的检索性能已经取得显著进步，但在多标签检索任务上仍有一定的提升空间。因此在下一步工作中，本文将考虑进一步提高本文方法对多标签数据的适应能力。

参考文献：

[1] Liu H M, Wang R P, Shan S G, *et al.* Deep supervised hashing for fast image retrieval [C]// Proc of the 2016 IEEE Conference on Computer Visi-on and Pattern Recognition. Washington: IEEE Computer Society, 2016: 2064-2072.

[2] Gong Y C and Lazebnik S. Iterative quantization: a procrustean approach to learning binary codes [C]// Proc of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2011: 817-824.

[3] Andoni A and Razenshteyn I. Optimal data dependent hashing for approximate near neighbors [C]// Proc of the 47th Annual ACM Symposium on the Theory of Computing. New York: ACM Press, 2015: 793-801.

[4] Andoni A and Indyk P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions [C]// Proc of the 2006 Annual IEEE Symposium on Foundations of Computer Science. Piscataway, NJ: IEEE Press, 2006: 459-468.

[5] Kong W H and Li W J. Isotropic hashing [C]// Proc of the 2012 Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1655-1663.

[6] Atkinson M P, Orlowska M E, Valduriez P, *et al.* Similarity search in high dimensions via hashing [C]// Proc of 25th International Conference on Very Large Data Bases. San Francisco: Morgan Kaufmann, 1999: 518-529.

[7] Liu W, Wang J, Ji R R, *et al.* Supervised hashing with kernels [C]// Proc of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2012: 2074-2081.

[8] Krizhevsky A, Sutskever I and Hinton G. ImageNet classification with deep convolutional neural networks [C]// Proc of the 2012 Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1106-1114.

[9] Szegedy C, Liu W, Jia Y Q, *et al.* Going deeper with convolutions [C]// Proc of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2015: 7-12.

[10] Sun Y, Chen Y H, Wang X G, *et al.* Deep learning face representation by joint identification-verification [C]// Proc of the 2014 Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 1988-1996.

[11] Taigman Y, Yang M, Ranzato M, *et al.* DeepFace: closing the gap to human-level performance in face verification [C]// Proc of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2014: 1701-1708.

[12] Zhao F, Huang Y Z, Wang L, *et al.* Deep semantic ranking based hashing for multi-label image retrieval [C]// Proc of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE

ChinaXiv合作期刊

chinaXiv:202205.00069v1

Computer Society, 2015: 1556-1664.

[13] Zhang R M, Lin L, Zhang R, *et al.* Bit-Scalable deep hashing with regularized similarity learning for image retrieval and person re-identification [J]. IEEE Trans on Image Processing, 2015, 24 (12): 4766-4779.

[14] Lai H J, Pan Y, Liu Y, *et al.* Simultaneous feature learning and hash coding with deep neural networks [C]// Proc of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2015: 3270-3278.

[15] Li Q, Sun Z N, He R, *et al.* Deep supervised discrete hashing [J]. IEEE Trans on Image Processing, 2018: 27 (12): 5996-6009.

[16] Luo X, Chen C, Zhang H S, *et al.* A Survey on Deep Hashing Methods [J]. Arxiv, 2020, abs/2003. 03369.

[17] Xia H K, Pan Y, Lai H J, *et al.* Supervised hashing for image retrieval via image representation learning [C]// Proc of the 28th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2014: 2156-2162.

[18] Li W J, Wang S and Kang W C. Feature learning based deep supervised hashing with pairwise Labels [C]// Proc of the 25th International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2016: 1711-1717.

[19] Zhu H, Long M S, Wang J M, *et al.* Deep hashing network for efficient similarity retrieval [C]// Proc of the 30th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2016: 2415-2421.

[20] Wang X F, Shi Y and Kitani K M. Deep supervised hashing with triplet labels [C]// Proceedings of the 13th Asian Conference on Computer Vision. Taipei, Taiwan: Springer, 2016: 70-84.

[21] Jiang Q Y and Li W J. Asymmetric deep supervised hashing [C]// Proc of the 32th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2018: 3342-3349.

[22] Dong W, Charikar M, Li K, Asymmetric distance estimation with sketches for similarity search in high-dimensional spaces [C]. Proc of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2008: 123-130.

[23] Gordo A, Perronnin F, Gong Y C, *et al.* Asymmetric distances for binary embeddings [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2014, 36 (1): 33-47.

[24] Shen F M, Gao X, Liu L, *et al.* Deep asymmetric pairwise hashing [C]. /Proc of the 2017 ACM on Multimedia Conference. New York: ACM Press, 2017: 1522-1530.

[25] Da C, Meng G F, Xiang S M, *et al.* Nonlinear asymmetric multi-valued hashing [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2019, 44 (11): 2660-2676.

[26] Luo Y D, Huang Z, Li Y, *et al.* Collaborative learning for extremely low bit asymmetric hashing [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2021, 33 (12): 3675-3685.

[27] Lu J L, Wang H L, Zhou J, *et al.* Deep Asymmetric Hashing with Dual Semantic Regression and Class Structure Quantization [J]. Information Sciences, 2022, 589: 235-249.

[28] 顾广华，霍文华，苏明月，等. 基于非对称监督深度离散哈希的图像检索 [J]. 电子与信息学报, 2021, 43 (12): 3530-3537. (Gu Guanghua, Huo Wenhua, Su Mingyue, *et al.* Asymmetric Supervised Deep Discrete Hashing Based Image Retrieval [J]. Journal of Electronics & Information Technology, 2021, 43 (12): 3530-3537.)

[29] Zhu H and Gao S H. Locality constrained deep supervised hashing for image retrieval [C]// Proc of the 26th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2017: 3567-3573.

[30] Yang H F, Lin K and Chen C S. Supervised learning of semantics-preserving hashing via deep convolutional neural networks for large-scale image search [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 40 (2): 437-451.

[31] Su S P, Zhang C, Han K, *et al.* Greedy Hash: towards fast optimization for accurate hash coding in CNN [C]// Proc of the 2018 Neural Information Processing Systems. Cambridge, MA: MIT Press, 2018: 806-815.

[32] Krizhevsky A and Hinton G. Learning multiple layers of features from tiny images [D]. Technical Report TR-2009. University of Toronto, 2009.

[33] Chua T S, Tang J H, Hong R C, *et al.* NUS-WIDE: a real-world web image database from national university of singapore [C]// Proc of the 8th ACM International Conference on Image and Video Retrieval. New York: ACM Press, 2009: 1-9.

[34] Kingma D P and Ba J. 2014. Adam: a method for stochastic optimization. [EB/OL]. (2020-02-25) . https://arxiv. org/pdf/1412. 6980. pdf.

[35] Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115 (3): 211-252.

[36] Neyshabur B, Srebro N, Salakhutdinov R, *et al.* The power of asymmetry in binary hashing [C]// Proc of the 2013 Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013: 2823-2831.

[37] Zhang P C, Zhang W, Li W J, *et al.* Supervised hashing with latent factor models [C]// Proc of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2014: 173-182.

[38] Lin G S, Shen C H, Shi Q F, *et al.* Fast supervised hashing with decision trees for high-dimensional data [C]// Proc of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2014: 1971-1978.

[39] Shen F M, Shen C H, Liu W, *et al.* Supervised discrete hashing [C]// Proc of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 37-45.

[40] Kang W C, Li W J and Zhou Z H. Column sampling based discrete supervised hashing [C]// Proc of the 30th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2016: 1230-1236.

[41] Shi X S, Xing F Y, Xu K D, *et al.* Asymmetric discrete graph hashing [C]// Proceedings of the 31th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2017: 2541-2547.

[42] 张志升，曲怀敬，徐佳，等. 稀疏差分网络和多监督哈希用于高效图像检索 [J]. 计算机应用研究, 2021, 39 (6): 1-8. (Zhang Zhisheng, Qu Huaijing, Xu Jia, *et al.* Sparse differential network and multi-supervised hashing for efficient image retrieval [J]. Application Research of Computers, 2021, 39 (6): 1-8.)

[43] Chen Y B, Zhang S, Liu F X, *et al.* TransHash: Transformer-based Hamming Hashing for Efficient Image Retrieval [J]. Arxiv, 2021, abs/2105. 01823.